University of Zurich[UZH]

Zurich Center for Neuroeconomics

Model-free or muddled models in the two-stage task?

Carolina Feher da Silva[1], Todd Hare[1]

[1] Zurich Center for Neuroeconomics, Department of Economics, University of Zurich

Introduction

Choice patterns in the two-stage task developed by Daw et al. (2011) suggest that the human brain employs both model-free and model-based learning. However, another possibility is that the apparent model-free influences are caused by model-based learning operating with an incorrect model of the task. We sought to test this hypothesis.

Methods

Human participants performed a two-stage task with storified instructions that gave a reason for rare transitions. The “magic carpet” and “spaceship” versions of the task (see below) included 21 and 24 participants, respectively. In both cases, the task had the same stage transition and reward probabilities (and hence the same tradeoff between accuracy and effort) as the original two-stage task in Daw et al. (2011).

We also simulated three types of purely model-based agents performing the two-stage task for 1000 trials:

1. **Original**: agents that use the assumed world model, for comparison.
2. **Unlucky symbol**: agents that incorrectly assume that choosing a particular first-stage symbol lowers the values of second-stage symbols by 50%.
3. **Transition-dependent learning rates (TDLR)**: agents that have a higher learning rate if the transition is common (learning rate = 0.8) versus rare (learning rate = 0.2).

We analyzed data from the two-stage tasks by logistic regression of consecutive trials and by fitting a hybrid model-based/model-free reinforcement learning model to the data.

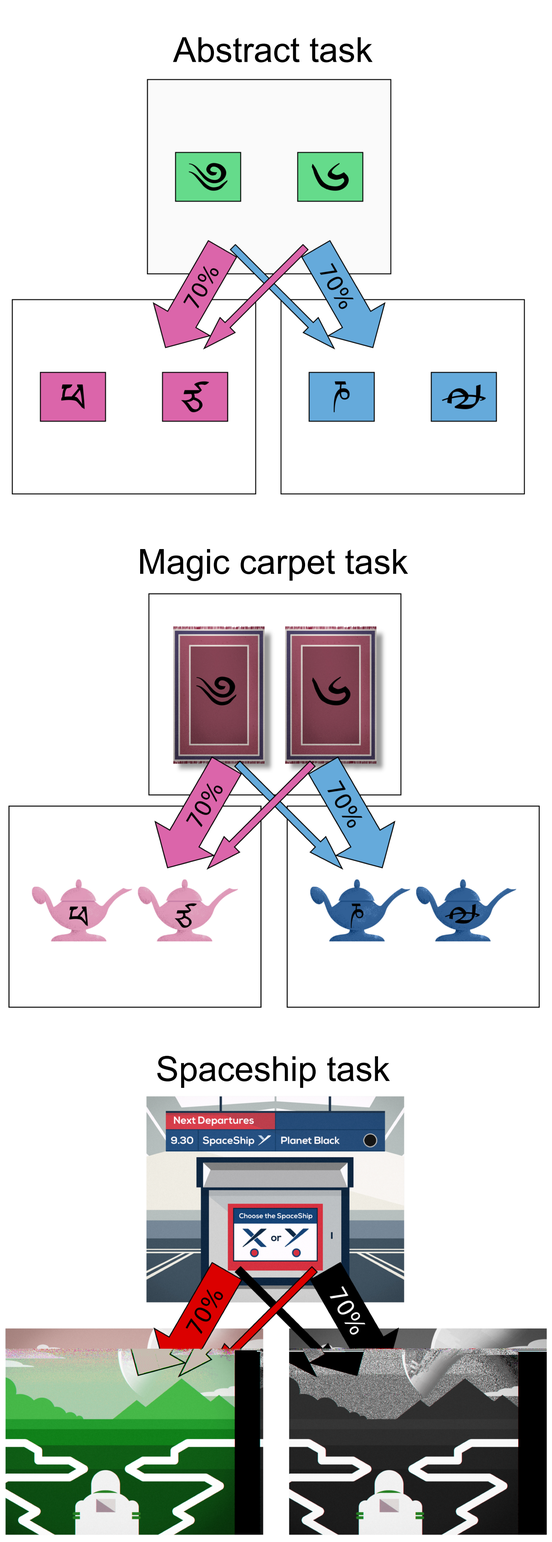

Fig.1: Transition model of two-stage tasks: each first-stage choice transitions to a different second-stage state with 0.7 probability and to the other second-stage state with 0.3 probability.

Simulation results

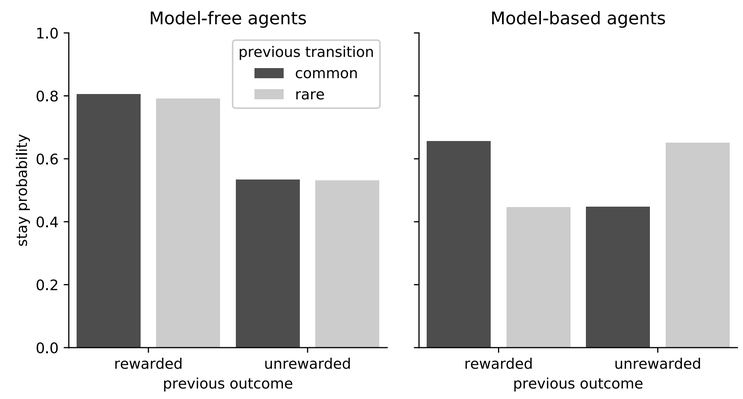

Fig.2: The canonical pattern of behavior in simulated model-free and model-based agents performing the two-stage task was obtained by logistic regression of consecutive trial pairs.

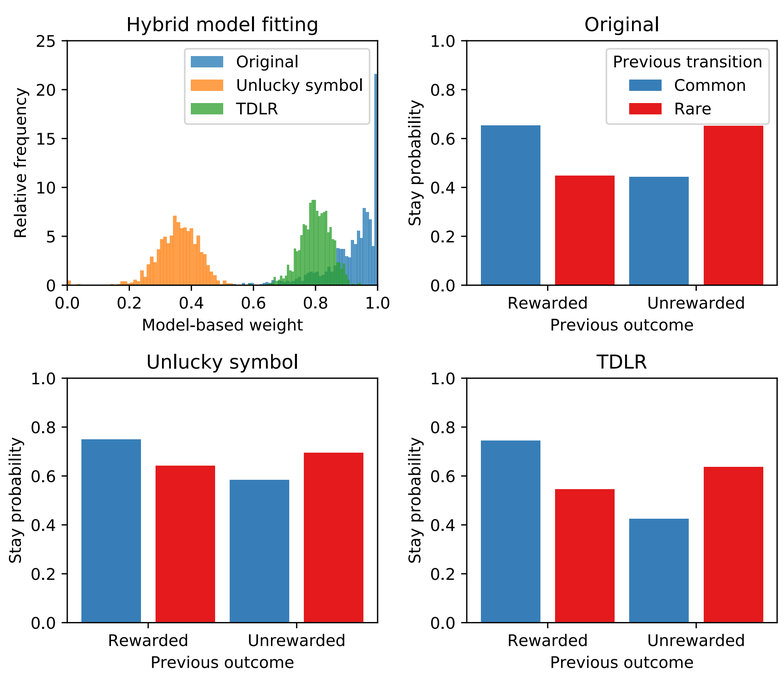

Fig.3: Purely model-based agents can be confused with model-free agents if the world model used by the agents breaks the analysis’s assumptions. Three types of simulated model-based agents are described in the Methods. Simulated choices were analyzed by fitting a hybrid reinforcement learning model (Daw et al., 2011) to the data (upper left plot) and by logistic regression of consecutive trial pairs (upper right and lower plots). These analyses *incorrectly* suggest that the purely model-based “unlucky symbol” and “transition-dependent learning rates (TDLR)” agents are hybrids between model-based and model-free influence. (N = 1000 per simulation.)

Human behavioral results

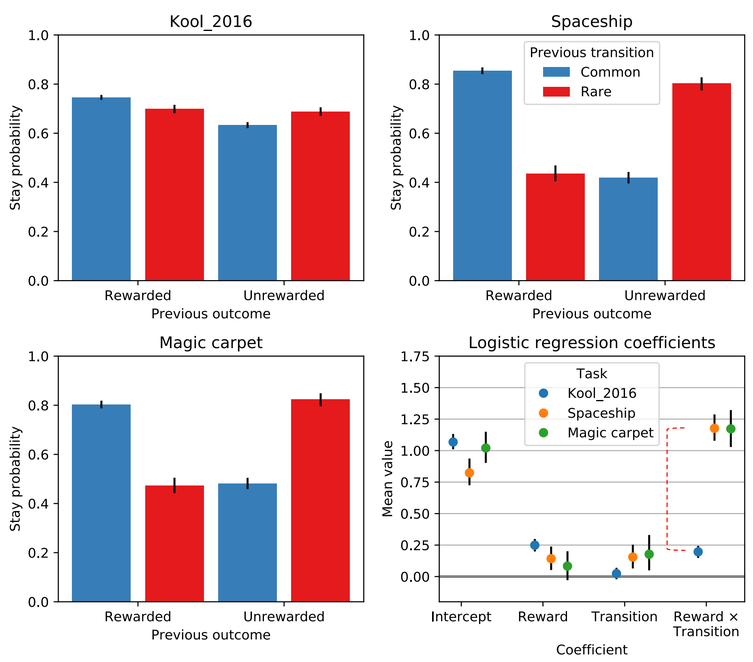

Fig.4: Logistic regression and hybrid model analysis of behavior on differently instructed versions of the task. The bar plots show that storified instructions lead to stronger reward by transition interactions. The coefficient for the reward by transition interaction was 5.9 times larger in both versions of our instructions (95% CI [4.7, 7.2]; 95% CI [4.5, 7.3]) than in Kool et al. (2016).

Funding

1. EU FP7 # 607310 (TAH)
2. CAPES # 88881.119317/2016-01 (CFS)

Nudge-it.

Conclusions

- Confusion about how the task works (i.e. employing incorrect models of the task structure) can lead to behavior that mimics model-free reinforcement learning.
- Explaining the two-stage task as a detailed story with reasons for the rare transitions leads to increased model-based behavior.

Task	25%	Median	75%
Original	0.29	0.39	0.59
Kool_2016	0.00	0.27	0.66
Magic carpet	0.56	0.76	0.84
Spaceship	0.51	0.79	0.85

Fig.5: Model-based weights estimated by maximum likelihood for participants in the original Daw et al.’s study (N = 17), Kool et al.’s replication (N = 206), magic carpet task (N = 21), and spaceship task (N = 24).

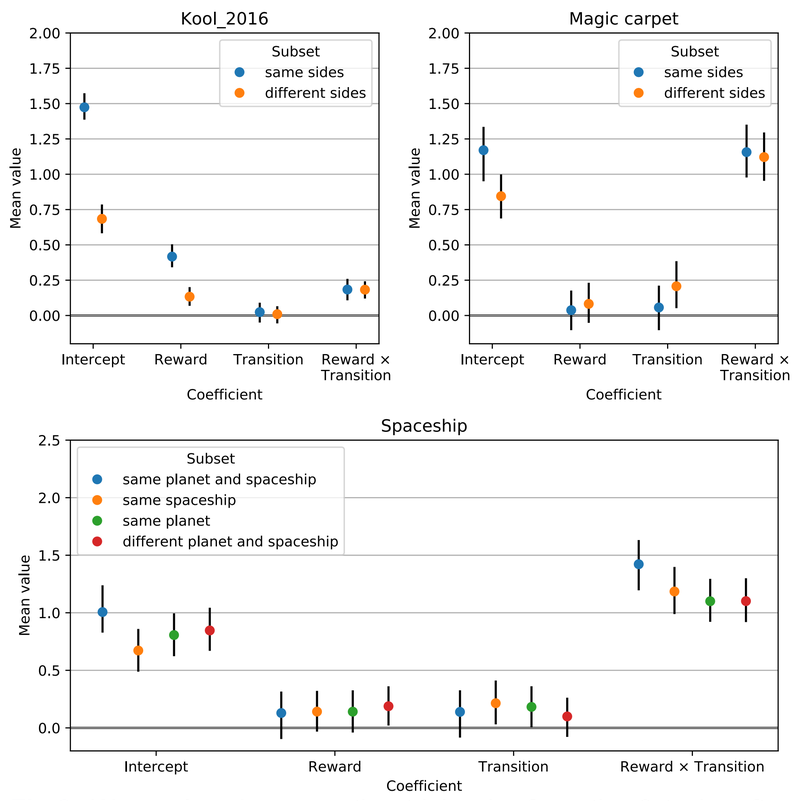

Fig.6: We analyzed consecutive trial pairs from the 3 tasks after separating the pairs by the identity of the first-stage stimuli. In Kool_2016, reward effects are present, but smaller when the initial stimuli were displayed on different sides of the screen. In the spaceship task, reward effects were similar across different initial stimuli that have the same meaning – contrary to the predictions of model-free driven reward effects. This suggests that the reward effect observed for the spaceship task was caused by an incorrect model of the task rather than model-free learning.

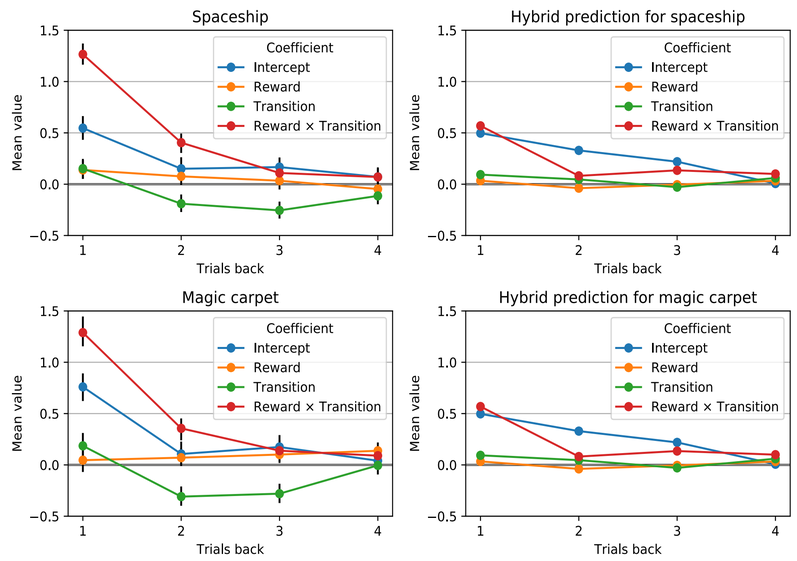

Fig.7: A traditional hybrid reinforcement learning model will not always accurately capture participants’ behavior on the two-stage task. We analyzed both our tasks using logistic regressions with the 4 previous trials as predictors. Plots in the left column show the behavioral results, and those on the right show patterns generated by the best-fitting hybrid-model parameters. There is substantial mismatch between the observed and predicted data.

References

1. Daw et al. *Model-Based Influences on Humans’ Choices and Striatal Prediction Errors*. Neuron 69, 1204–1215, 2011.
2. Kool et al. *When Does Model-Based Control Pay Off?* PLoS Comput Biol 12(8): e1005090, 2016.